CLAIMS

1. A speech encoder, comprising:

a content extraction module including,

a band pass filter that receives a speech input signal and generates a band limited speech signal,

a first speech buffer connected to the band pass filter that stores the band limited speech signal,

an LP analysis block connected to the first speech buffer that reads the stored speech signal and generates a plurality of LP coefficients therefrom,

an LPC to LSF block connected to the LP analysis block for converting the LP coefficients to a line spectral frequency (LSF) vector,

an LP analysis filter connected to the LPC to LSF block that extracts an LP residual signal from the LSF vector; and

an LSF quantizer connected to the LPC to LSF block that receives the LSF vector and determines an LSF index therefor;

a pitch detector connected to the LP analysis block of the content extraction module, the pitch detector classifying the band filtered speech signal as one of a voiced signal and an unvoiced signal; and

a naturalness enhancement module connected to the content extraction module and the pitch detector, the naturalness enhancement module including,

means for extracting parameters from the LP residual signal, wherein for an unvoiced signal the extracted parameters include pitch and gain and for a

voiced signal the extracted parameters include pitch, gain and excitation level; and

a quantizer for quantizing the extracted parameters and generating quantized parameters.

2. The speech encoder of claim 1, wherein the band pass filter comprises an eighth order IIR filter.

3. The speech encoder of claim 3, wherein the IIR filter includes a fourth order low-pass section and a fourth order high pass section.

4. The speech encoder of claim 1, further comprising a scale down unit connected between the band pass filter and the first speech buffer, wherein the scale down unit limits a dynamic range of the band limited speech signal and provides a scaled down signal to the first speech buffer.

5. The speech encoder of claim 4, wherein the scale down unit scales the band limited speech signal by about 0.5.

6. The speech encoder of claim 1, wherein the LP analysis block performs a 10^{th} order Burg's LP analysis to estimate a spectral envelope of the stored speech signal and generate the plurality of LP coefficients.

7. The speech encoder of claim 7, wherein a bandwidth expansion block expands the plurality of LP coefficients to generate bandwidth expanded LP coefficients.

8. The speech encoder of claim 1, wherein the naturalness enhancement module uses different update rates to extract each parameter.

9. The speech encoder of claim 8, wherein the update rate of the gain is about 5 mS and the update rates of the pitch frequency and excitation level are about 10 mS.

10. The speech encoder of claim 1, wherein the content extraction module further includes a first residual buffer for storing the LP residual signal.

11. The speech encoder of claim 10, wherein the parameters are extracted from the LP residual signal stored in the first residual buffer.

12. The speech encoder of claim 1, wherein for an unvoiced signal, the pitch parameter is set to zero to distinguish the unvoiced signal pitch from the voiced signal pitch.

13. The speech encoder of claim 1, wherein the naturalness enhancement module further includes a down-sampler connected between the parameter extraction means and the quantizer, for down sampling the parameters prior to quantization.

14. The speech encoder of claim 13, wherein the pitch and excitation parameters are downsampled at a rate of about 4:1.

15. The speech encoder of claim 13, wherein the pitch and excitation parameters are downsampled at a rate of about 2:1.

16. The speech encoder of claim 1, wherein the pitch detector distinguishes between an unvoiced signal and a voiced signal using an RMS value and an energy distribution of the scaled-down, band-filtered speech signal.

17. The speech encoder of claim 1, wherein the pitch detector has three levels of operation depending on an ambiguity level of the scaled-down, band-filtered speech signal.

18. The speech encoder of claim 17, wherein the first level of operation of the pitch detector includes:

a low pass filter that receives the scaled-down, band-filtered speech signal and rejects a high frequency content thereof;

a second speech buffer connected to the low pass filter for storing the low pass filtered signal;

an inverse filter connected to the second speech buffer for generating a band-limited residual signal from the low pass filtered signal stored in the second speech buffer;

a cross-correlation function generator, connected to the inverse filter, for generating a cross-correlation function of the band-limited residual signal;

a peak detector, connected to the cross-correlation function generator, for detecting a global maximum across the cross-correlation function and a location of the global maximum;

a level detector connected to the peak detector for comparing the cross-correlation function global maximum to a predetermined value and based on the comparison result, classifying the input speech signal as one of a voiced signal and an unvoiced signal; and

means for generating a first estimated pitch period based on the cross-correlation function.

19. The speech encoder of claim 18, wherein the second level of operation of the pitch detector includes:

means for computing an RMS value of the speech signal;

means for computing an energy distribution of the speech signal; and

means for comparing the computed RMS value and the computed energy distribution with first and second cut-off values to determine whether the speech signal is a voiced or unvoiced signal, wherein if the result of the comparison indicates that the speech signal is an unvoiced signal, then the second estimated pitch period is set to zero.

20. The speech encoder of claim 18, wherein the third operation level includes:

means for eliminating multiple pitch errors, connected to the level detector, the multiple pitch error elimination means generating the third estimated pitch period.

21. The speech encoder of claim 18, wherein a cutoff frequency of the low pass filter is about 1000Hz.

22. A content extraction module for a speech encoder, the content extraction module comprising:

a band pass filter that receives a speech input signal and generates a band limited speech signal,

a first speech buffer connected to the band pass filter that stores the band limited speech signal,

an LP analysis block connected to the first speech buffer that reads the stored speech signal and generates a plurality of LP coefficients therefrom,

an LPC to LSF block connected to the LP analysis block for converting the LP coefficients to a line spectral frequency (LSF) vector,

an LP analysis filter connected to the LPC to LSF block that extracts an LP residual signal from the LSF vector; and

an LSF quantizer connected to the LPC to LSF block that receives the LSF vector and determines an LSF index therefor.

23. The content extraction module of claim 22, wherein the band pass filter comprises an eighth order IIR filter.

24. The content extraction module of claim 23, wherein the IIR filter includes a fourth order low-pass section and a fourth order high pass section.

25. The content extraction module of claim 22, further comprising a scale down unit connected between the band pass filter and the first speech buffer, wherein the scale down unit limits a dynamic range of the band limited speech signal and provides a scaled down signal to the first speech buffer.

26. The content extraction module of claim 25, wherein the scale down unit scales the band limited speech signal by about 0.5.

27. The content extraction module of claim 22, wherein the LP analysis block performs a 10^{th} order Burg's LP analysis to estimate a spectral envelope of the stored speech signal and generate the plurality of LP coefficients.

28. The content extraction module of claim 27, wherein a bandwidth expansion block expands the plurality of LP coefficients to generate bandwidth expanded LP coefficients.

29. The content extraction module of claim 22, further comprising a first residual buffer for storing the LP residual signal.

30. A naturalness enhancement module for a speech encoder, wherein the speech encoder includes a pitch detector for determining whether an input speech signal is a voiced signal or an unvoiced signal and a content extraction module for generating an LP residual signal from

the input speech signal, the naturalness enhancement module comprising:

means for extracting parameters from the LP residual signal, wherein for an unvoiced signal the extracted parameters include pitch and gain and for a voiced signal the extracted parameters include pitch, gain and excitation level; and

a quantizer for quantizing the extracted parameters and generating quantized parameters.

31. The naturalness enhancement module of claim 30, wherein the naturalness enhancement module uses different update rates to extract the parameters from the LP residual signal.

32. The naturalness enhancement module of claim 31, wherein the update rate of the gain is about 5 mS and the update rates of the pitch frequency and excitation level are about 10 mS.

33. The naturalness enhancement module of claim 31, wherein for an unvoiced signal, the pitch parameter is set to zero to distinguish the unvoiced signal pitch from the voiced signal pitch.

34. The naturalness enhancement module of claim 33, further comprising a down-sampler connected between the parameter extraction means and the quantizer, for down sampling the parameters prior to quantization.

35. The naturalness enhancement module of claim 34, wherein the pitch and excitation parameters are downsampled at a rate of about 4:1.

36. The naturalness enhancement module of claim 33, wherein the pitch and excitation parameters are downsampled at a rate of about 2:1.

37. A pitch detector for a speech encoder, the pitch detector comprising:

a first operation level for analyzing a speech signal and, based on a first predetermined ambiguity value of the speech signal, generating a first estimated pitch period; and

a second operation level for analyzing the speech signal and, based on a second predetermined ambiguity value of the speech signal, generating a second estimated pitch period.

38. The pitch detector of claim 37, further comprising:

a third operation level for analyzing the speech signal and, based on a third ambiguity level of the speech signal, generating a third estimated pitch period.

39. The pitch detector of claim 38, wherein the first operation level includes:

a low pass filter that receives the speech signal and rejects a high frequency content thereof;

a speech buffer connected to the low pass filter for storing the low pass filtered speech signal;

an inverse filter connected to the speech buffer for generating a residual signal from the low pass filtered speech signal stored in the second speech buffer;

a residual buffer connected to the inverse filter for storing the residual signal;

a first cross-correlation function generator, connected to the residual buffer, for generating a first cross-correlation function of the residual signal stored in the residual buffer;

a peak detector, connected to the cross-correlation function generation means, for detecting a global maximum across the cross-correlation function and a location of the global maximum; and

a level detector, connected to the peak detector, for comparing the cross-correlation function global maximum to the first predetermined ambiguity value and to classify the input speech signal as a voiced signal or an unvoiced signal in response to the comparison; and

means for calculating the first estimated pitch period based on the cross-correlation function.

40. The pitch detector of claim 39 wherein if the global maximum is less than the predetermined ambiguity level than the speech signal is classified as an unvoiced signal.

41. The pitch detector of claim 39 wherein a cutoff frequency of the low pass filter is about 1000Hz.

42. The pitch detector of claim 39, wherein the second operation level includes:

means for computing an RMS value of the speech signal;

means for computing an energy distribution of the speech signal; and

means for comparing the computed RMS value and the computed energy distribution with first and second cut-off values to determine whether the speech signal is a voiced or unvoiced signal, wherein if the result of the comparison indicates that the speech signal is an unvoiced signal, then the second estimated pitch period is set to zero.

43. The pitch detector of claim 42, wherein the third operation level includes:

means for eliminating multiple pitch errors, connected to the level detector, the multiple pitch error elimination means generating the third estimated pitch period.

44. A speech signal preprocessor for preprocessing an input speech signal prior to providing said speech signal to a speech encoder, the preprocessor comprising:

a band pass filter that receives said speech input signal and generates a band limited speech signal; and

a scale down unit connected to the band pass filter for limiting a dynamic range of the band limited speech signal.

45. The speech signal preprocessor of claim 44, wherein the band pass filter comprises an eighth order IIR filter.

46. The speech signal preprocessor of claim 45, wherein the IIR filter includes a fourth order low-pass section and a fourth order high pass section.

47. The speech signal preprocessor of claim 44, wherein the scale down unit scales the band limited speech signal by about 0.5.

48. A method of encoding a speech signal, comprising the steps of:

filtering the speech signal to limit a bandwidth thereof;

fragmenting the filtered speech signal into speech segments;

decomposing the speech segments into a spectral envelope and an LP residual signal, wherein the spectral envelope is represented by a plurality of LP filter coefficients (LPC);

converting the LPC into a plurality of line spectral frequencies (LSF);

classifying each speech segment as one of a voiced segment and an unvoiced segment based on a pitch of the segment;

extracting parameters from the LP residual signal, wherein for an unvoiced segment the extracted parameters include pitch and gain and for a voiced segment the extracted parameters include pitch, gain and excitation level; and

quantizing the extracted parameters and generating quantized parameters.

49. The method of encoding a speech signal of claim 48, wherein the speech signal is filtered with an eighth order IIR filter.

50. The method of encoding a speech signal of claim 49, wherein the IIR filter includes a fourth order low-pass section and a fourth order high pass section.

51. The method of encoding a speech signal of claim 48, further comprising the step of scaling the filtered speech signal prior to the fragmenting step.

52. The method of encoding a speech signal of claim 49, wherein the decomposing step performs a 10^{th} order Burg's LP analysis to estimate the spectral envelope of the speech segments and generate the LP filter coefficients.

53. The method of encoding a speech signal of claim 49, wherein the extracting parameters step uses different update rates to extract each parameter.